MULTIMEDIA UNIVERSITY

STUDENT IDENTIFICATION NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 2, 2016/2017

BST1034 – Statistics for Managers

(All sections / Groups)

2 MARCH 2017
2.30 p.m. – 4.30 p.m.
(2 Hours)

INSTRUCTIONS TO STUDENT

1. This Question paper consists of 9 pages with 4 Questions only.
2. Attempt **all FOUR** questions. All questions carry equal marks and the distribution of the marks for each question is given.
3. Students are allowed to use scientific calculators (without restriction).
4. Statistical formulae are attached at the end of the question paper.
5. Please write all your answers in the Answer Booklet provided.

QUESTION 1

(a) Consider the following set of sample data:

269	159	254	154	148	156	136	199	148	300
255	249	163	147	195	199	198	201	211	189

(i) Compute the mean and variance for the sample data. (4 marks)

(ii) Calculate the coefficient of variation for the sample data. (2 marks)

(b) Of the travellers arriving at a small airport, 60% fly on major airlines, 30% fly on privately owned planes, and the remainder fly on commercially owned planes not belonging to a major airline. Of those travelling on major airlines, 50% are travelling for business reasons, whereas 60% of those arriving on private planes and 90% of those arriving on other commercially owned planes are travelling for business reasons. Suppose that we randomly select one person arriving at this airport.

(i) What is the probability that the person is travelling for business reason? (3 marks)

(ii) What is the probability that the person arrived on a privately-owned plane, given that the person is travelling for business reasons? (3 marks)

(c) Tay-Sachs disease is a genetic disorder that is usually fatal in young children. If both parents are carriers of the disease, the probability that their offspring will develop the disease is approximately 25%. Suppose that a husband and wife are both carriers and that they have three children. If the outcomes of the three pregnancies are mutually independent, what are the probabilities of the following events:

(i) All three children develop Tay-Sachs. (2 marks)

(ii) Only one child develops Tay - Sachs. (2 marks)

(d) A life insurance salesman sells on the average 3 life insurance policies per week.

(i) Calculate the probability that in a given week he will sell 2 or more policies but less than 5 policies. (3 marks)

(ii) Assuming that there are 5 working days per week, what is the probability that in a given day he will sell one policy? (3 marks)

(e) Mike have a small portfolio of shares in Companies A, B and C as listed in the table below:

	Year 1999		Year 2000	
	Price (RM)	Quantity	Price (RM)	Quantity
Company A	0.9	100	1	200
Company B	1	200	0.9	100
Company C	2	200	4	300

Calculate the Laspeyres Price Index for year 2000 for each of the following with base year 1999. (3 marks)

(Total: 25 Marks)

QUESTION 2

(a) In 2015, the mean score of basketball players in XYZ basketball competition is 8.5 points per game. In 2016, a random sample of 100 basketball players participated in XYZ basketball competition has a mean score of 9.7 points per game with a population standard deviation of 2.0 points per game.

(i) Construct a 95% confidence interval for the mean score per game of NBA players in year 2016. (6 marks)

(ii) At the 0.05 level of significance, test if the mean score per game of the basketball players participated in XYZ basketball competition has increased. (8 marks)

(b) There are two categories in XYZ basketball competition, U18 and U16. In 2016, a random sample of 10 basketball players are selected from U18 category and 20 basketball players are selected from U16 category. The performance of each category is listed in table below.

	U18	**U16**
Mean score per game	10 points	9.2 points
Standard deviation	1.5 points	2.25points

At the significance level of 0.01, perform a test of hypothesis to determine if the mean score per game of U18 basketball players is higher than U16 basketball players. (Assume equal variances) (11 marks)

(Total: 25 Marks)

Continued...

QUESTION 3

(a) A statistician employed by a television rating service wanted to determine if there were differences in television viewing habits among three different states in Malaysia. She took a random sample of five adults in each of the states and asked each to report the number of hours spent watching television in the previous week. The summary of the results is shown below. (Assume normal distributions with equal variances.)

SUMMARY

Groups	Count	Sum	Average	Variance
Selangor	5	124	24.8	23.2
Penang	5	161	32.2	12.7
Johor Bahru	5	94	18.8	10.2

ANOVA

Source of Variation	SS	df	MS	F
Between Groups	***	***	***	***
Within Groups	***	12	***	
Total	634.933	14		

(i) Complete the ANOVA table. (7 marks)

(ii) Can she infer at the 5% significance level that differences in hours of television watching exist among the three states? (6 marks)

(b) An article reports that smaller firms seems to be hiring more than large ones as the company picks up its pace. The table below gives the number of employees hired and those laid off, out of a random sample of 1032, broken down by firm size.

	Small Firm	Medium Size Firm	Large Firm
Number hired	210	290	325
Number laid off	32	95	80

Is there evidence that hiring practices are dependent on firm size? Use $\alpha = 0.025$. (12 marks)

(Total: 25 marks)

Continued...

QUESTION 4

(a) Study has shown that the price of gold is highly influenced by the U.S. interest rates. The following data shows the yearly averages of U.S. real interest rates (in percentage) and prices of gold per ounce (in thousand USD) in the ten most recent years.

U.S. real interest rate	4.74	5.25	3.07	2.47	2.00	1.16	1.38	1.59	1.58	2.24
Gold price	0.60	0.70	0.87	0.97	1.22	1.57	1.67	1.41	1.27	1.16

(i) Compute the coefficient of correlation. (3 marks)

(ii) Find the regression line of the gold price. (5 marks)

(b) The following are the partially completed outputs of the profitability of a firm regressed on four main factors, that is firm size, productivity, competition, and economy condition:

Source of Variation	*SS*	*df*	*MS*
Regression	***	4	***
Residual	9.782	***	***
Total	27.646	138	

Variable	*t Stat*
Firm size	2.545
Productivity	2.757
Competition	-1.690
Economy condition	3.001

(i) Complete the ANOVA table. (3 marks)

(ii) Test whether the regression model is statistically significant at 10% significance level. (8 marks)

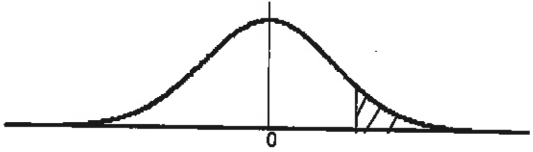
(iii) Which variable is significant at 5% significance level? Would you consider omitting any variable? (6 marks)

(Total: 25 Marks)

End of Paper

Table 1

The Upper Tail Area Under the Standard Normal Curve



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4.0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

STATISTICAL FORMULAE

A. DESCRIPTIVE STATISTICS

Ungrouped Data

Mean $= \frac{\sum x}{n}$

Standard Deviation (s) $= \sqrt{\frac{\sum x^2}{n-1} - \frac{(\sum x)^2}{n(n-1)}}$

Grouped Data

Mean $= \frac{\sum fx}{\sum f}$

Standard Deviation (s) $= \sqrt{\frac{\sum fx^2}{\sum f - 1} - \frac{(\sum fx)^2}{\sum f(\sum f - 1)}}$

Median $= L_m + \left[\frac{\left(\frac{\sum f}{2} - F_{m-1}\right)}{f_m}\right]c$

Mode $= L_m + \left[\frac{(f_m - f_{m-1})}{2f_m - f_{m-1} - f_{m+1}}\right]c$

Coefficient of Variation $= \frac{s}{x} \times 100\%$

B. PROBABILITY

Poisson Probability Distribution

If X follows a Poisson Distribution P (λ) where $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$

then the mean = E(X) = λ and variance = VAR(X) = λ

Binomial Probability Distribution

If X follows a Binomial Distribution B(n, p) where $P(X = x) = {}^nC_x p^x q^{n-x}$

then the mean = E(X) = n p and variance = VAR(X) = npq where q = 1 - p

Normal Distribution

If X follows a Normal distribution N(μ , σ) where E(X) = μ and VAR(X) = σ^2

then $Z = \frac{X - \mu}{\sigma}$

C. CONFIDENCE INTERVAL ESTIMATION AND SAMPLE SIZE DETERMINATION

(100 - α)% Confidence Interval for Population Mean,(σ Known) is $\mu = \bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ (100 - α)% Confidence Interval for Population Mean (σ Unknown) is $\mu = \bar{x} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}$
(100 - α)% Confidence Interval for Population Proportion is $\pi = p \pm Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$
Sample Size Determination for Population Mean = $n = \left(\frac{Z_{\alpha/2}\sigma}{e}\right)^2$ Sample Size Determination for Population Proportion = $n = \frac{(Z_{\alpha/2})^2\pi(1-\pi)}{e^2}$ Where e = Limit of Error in Estimation

D. HYPOTHESIS TESTING

One Sample Mean Test	
***Z*-test**	***t*-test**
$Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$	$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$
One Sample Proportion Test	
$Z = \frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$	

Two Sample Mean Test

***Z*-test**

$$Z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^{\,2}}{n_1} + \frac{\sigma_2^{\,2}}{n_2}}}$$

***t*-test**

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{where} \quad S_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Two Sample Proportion Test

$$Z = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\overline{p}(1 - \overline{p})\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} \quad \text{where} \quad \overline{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

where x_1 and x_2 are the number of successes from each population

E. ANALYSIS OF VARIANCE

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_j}(X_{ij} - \overline{\overline{X}})^2 \qquad SSB = \sum_{i=1}^{k} n_i(\overline{X}_i - \overline{\overline{X}})^2$$

$$SSW = \sum_{i=1}^{k}\sum_{j=1}^{n_j}(x_{ij} - \overline{x}_i)^2 \quad \text{or} \quad SSW = \sum_{i=1}^{k}(n_i - 1)s_i^2$$

F. CHI-SQUARE ANALYSIS

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i}$$

o_i = observed frequency in a particular cell
e_i = expected frequency in a particular cell

***Expected cell frequencies for Independent Test*:**

$$e_i = \frac{\text{row total} \times \text{column total}}{n}$$

G. REGRESSION ANALYSIS

Simple Linear Regression

$$Y = a + bX, \quad \text{where } b = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} \quad \text{and } a = \bar{y} - b\bar{x}$$

Correlation Coefficient

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

H. INDEX NUMBERS

Simple Price Index $P = \frac{p_t}{p_0} \times 100$	**Laspeyres Quantity Index** $P = \frac{\sum p_0 q_t}{\sum p_0 q_0} \times 100$
Simple Aggregate Price Index $P = \frac{\sum p_t}{\sum p_0} \times 100$	**Paasche Quantity Index** $P = \frac{\sum p_t q_t}{\sum p_t q_0} \times 100$
Laspeyres Price Index $P = \frac{\sum p_t q_0}{\sum p_0 q_0} \times 100$	**Fisher's Ideal Price Index** $\sqrt{(\text{Laspeyres Price Index})(\text{Paasche Price Index})}$
Paasche Price Index $P = \frac{\sum p_t q_t}{\sum p_0 q_t} \times 100$	**Value Index** $V = \frac{\sum p_t q_t}{\sum p_0 q_0} \times 100$